Journal of Signal Processing, Vol. 13, No. 4, pp. 339-342, July 2009

SELECTED PAPER

An MTF-Based Method of Blind Restoration for Improving Intelligibility of Bone-Conducted Speech

Kota Kinugasa, Masashi Unoki and Masato Akagi

School of Information Science, Japan Advanced Institute of Science and Technology
1-1 Asahidai, Tatsunokuchi, Nomi, Ishikawa, 923–1292, Japan
E-mail: {k-kota, unoki, akagi}@jaist.ac.jp

Abstract

Bone-conducted (BC) speech is useful for speech communications in extremely noisy environments. However, both the sound quality and intelligibility of BC speech are very poor. We propose a method of blind restoration for improving BC speech based on the concept of the modulation transfer function (MTF). We investigated the relationship between the power envelope of air-conducted (AC) and BC speech signals by analyzing an AC/BC database. We then modeled these relations by fitting some models to the MTF derived from the database. The best-fitted model had two parameters. The first was the gain factor and the second was the attenuation factor. We also propose a method of determining the parameters of the model without AC speech. We evaluated the new method using SNR, correlation, MTF, PESQ, and LSD. The results revealed that the proposed approach could effectively and blindly restore BC speech.

1. Introduction

It is very difficult for us to communicate through speech in extremely noisy environments such as those in factories and disaster sites. One of the best solutions is to use speech with a bone conduction microphone because BC speech can be recorded with this without interference from external noises. However, both the sound quality and intelligibility of BC speech are very poor [1]. Therefore, we need to compensate for these losses in BC speech to accomplish communication and this is a concern that presents numerous challenges.

The attenuation of BC speech is stronger than that of AC speech at higher frequencies. A straightforward method of restoring BC speech is to compensate for these attenuated frequency components by using high-pass filtering. Since attenuation in BC speech signals varies and can be complex depending on the BC pickup points, speakers, and pronounced syllables, it is very difficult to design one unique type of high-pass filtering with these variations. There are various methods of deriving inverse filtering such as cross-spectrum [2] and long-term Fourier transform methods [3]; however, these yield the restored speech signals with artifacts, i.e., echoes, so there are only slight improvements in voice quality.

However, by considering the relationship between AC and BC speech signals as a transfer function, we studied a common strategy based on the source-filter model to improve the intelligibility and sound quality of BC speech. As a result, we found that filter characteristics are more important than source characteristics to restore BC speech based on the source-filter model [4, 5]. Vu *et al.* proposed an LP-based method of blindly restoring BC speech that originated from the idea of the source-filter model in the frequency domain [5]. Although this could restore BC speech blindly, machine-learning methods had to be used to predict the AC-LP coefficients from BC-LP coefficients.

In contrast, Kimura *et al.* proposed a method of restoring BC speech based on the MTF concept in the time domain [4]. This method compensated for the reduced modulation index of all temporal power envelopes in the filterbank model. Because MTF is related to speech intelligibility, this method could directly improve the intelligibility of BC speech. However, AC speech was needed to restore BC speech with their method. Although it was extremely effective in restoring BC speech, it is debatable what type of model of MTF is the most useful for restoring BC speech.

Since it is very important to improve the loss of intelligibility as well as the voice quality of BC speech for speech communications, we propose an MTF-based method of blindly restoring BC speech. Thus, we attempted to analyze the characteristics of AC and BC power envelopes to design an MTF-based inverse filter and to find a method of blindly determining the parameters of the inverse filter without AC speech.

2. MTF-Based Model

The MTF concept was proposed by Houtgast and Steeneken [6] to predict speech intelligibility. Drullman revealed [7] that temporal envelope appears to be more important in speech intelligibility than carrier information. We think that the differences between AC and BC envelopes significantly affect speech intelligibility and sound quality.

Our method of restoring BC speech involves a filterbank, which is outlined in Fig. 1. Let $x(t)$ be AC speech and $y(t)$ be associated BC speech. We assume that the signals with the N-channel bandpass filterbank can be represented as

$$x(t) = \sum_{n=1}^{N} x_n(t) = \sum_{n=1}^{N} e_{x_n}(t) \cdot c_{x_n}(t) \quad (1)$$

$$y(t) = \sum_{n=1}^{N} y_n(t) = \sum_{n=1}^{N} e_{y_n}(t) \cdot c_{y_n}(t) \quad (2)$$

where $x_n(t)$ and $y_n(t)$ are the bandpass signals, $e_x(t)$ and $e_y(t)$ are the temporal envelopes, and $c_x(t)$ and $c_y(t)$ are

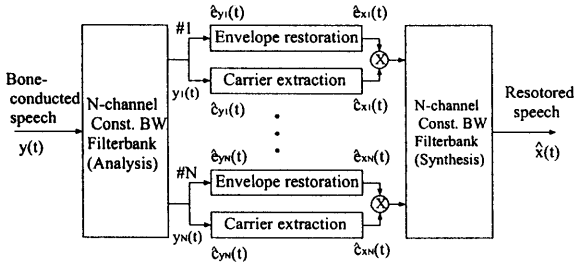

Figure 1: Method of restoration based on MTF

the carriers in the n-th channel of the filterbank at a constant bandwidth (40-Hz).

$$e^2_{y_n}(t) = \mathrm{LPF}\left[|y_n(t) + j\mathrm{Hilbert}(y_n(t))|^2\right] \quad (3)$$

$$c_{y_n}(t) = y_n(t)/e_{y_n}(t) \quad (4)$$

where Hilbert$(\cdot)$ is the Hilbert transform and LPF$[\cdot]$ denotes low-pass filtering at a 20-Hz cut-off frequency. $e_x(t)$ and $c_x(t)$ can also be calculated from $x(t)$ using the same method. Then, $E_h^{-1}(z)$ is used to restore the BC speech as

$$E_h^{-1}(z) = E_x(z)/E_y(z) \quad (5)$$

where $E_h(z)$, $E_x(z)$, and $E_y(z)$ correspond to the z-transform of $e^2_h(t)$, $e^2_x(t)$, and $e^2_y(t)$. Relation between $e^2_x(t)$ and $e^2_y(t)$, $e^2_h(t)$, was defined in our previous paper [4] as $e^2_h(t) = a^2_n \exp(-2b_n t)$, where a_n is the gain factor and b_n is the factor to control attenuation. However, there is no evidence whether the MTF model as $a^2_n \exp(-2b_n t)$ is the best representation for restoring BC speech.

3. Characteristics of Bone Conduction

The question is how can we design an inverse filter to restore BC speech? We analyzed the characteristics of bone conduction to design an MTF-based inverse filter, $E_h^{-1}(z)$.

3.1 AC/BC speech database

We used an AC/BC speech database [4] to analyze the characteristics of AC and BC power envelopes. BC speech was collected at five measurement points, i.e., (1: mandibular angle, 2: temple, 3: philtrum, 4: forehead, and 5: calvaria). Different microphones were used at points 1 to 4 and at point 5. Twenty-five Japanese words of each degree of familiarity were chosen from an NTT-database [8]. The speakers were five males and five females.

3.2 Analysis of characteristics of bone conduction

We analyzed the characteristics of AC and BC power envelopes using five measures: (1) the correlation between $e^2_x(t)$ and $e^2_y(t)$ (2) the SNR (S: $e^2_x(t)$, N: $e^2_x(t) - e^2_y(t)$), (3) the MTF, $M(\omega) = \left|\int_0^\infty e^2_h(t)\exp(-j\omega t)dt / \int_0^\infty e^2_h(t)dt\right|$,

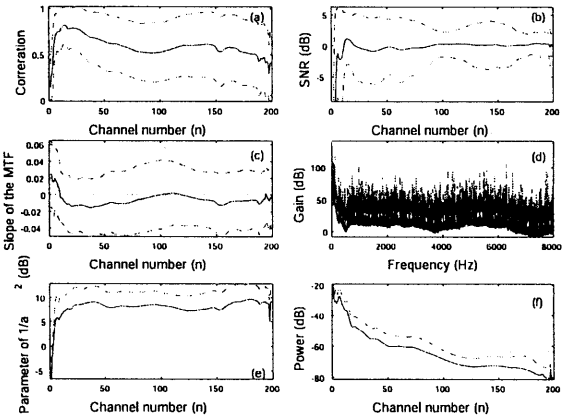

Figure 2: Results of analysis for all datasets (solid line: mean and dashed line: mean ± standard deviation): (a) correlation, (b) SNR, (c) slope of MTF, (d) transfer function, (e) mean of power ratio of power envelope ($1/a^2_n$) and regression curve, and (f) mean of $e^2_y(t)$ for each channel

(4) the transfer function, $|\mathcal{F}[y(t)]/\mathcal{F}[x(t)]|$, and (5) the power ratio, $a^2_n = 10\log_{10}(\int_0^T e^2_{y_n}(t)dt / \int_0^T e^2_{x_n}(t)dt)$. Here $\mathcal{F}[\cdot]$ is the long-term Fourier transform. Kimura *et al.* had analyzed the characteristics for measurement at point 5 (calvaria). We carefully analyzed and investigated the characteristics for measurements at all points.

3.3 Results and discussion

Figure 2 presents the results for all datasets and Fig. 3 has the results at measurement point 2. The solid lines indicate the mean and the dashed lines are the mean ± standard deviation. Figures 2(a) and (b) show the distortion in the BC power envelope. Figure 2(c) shows the slope of the regression line of MTF (1 to 10 Hz). The reason we limited the range of MTF from 1 to 10 Hz was that MTF at higher modulation frequencies is influenced by internal noise such as blood flow, transmission-line noise, and noise flooring. If the value of the slope is negative, MTF has low-pass characteristics and if the value of the slope is positive, MTF has high-pass characteristics. This indicates MTF has low-pass characteristics in most channels. Figure 2(d) indicates that bone conduction has low-pass characteristics. Figure 3(e) shows the mean for the power ratio of the power envelope of BC signals to AC signals in all channels. We fitted various curves to the mean of the power ratio and investigated whether the power ratio could be approximated by a regression curve as $1/\hat{a}^2_n = -cn^{-1} + d$, where c and d are parameters that depend on the measurement point. These characteristics were also in the results at the other measurement points in our analysis. We could approximate the power ratio of the BC power envelope to the AC power envelope for all channels by changing the parameters of the regression curve. Also, the results obtained from analysis reveal that the regression curve did not greatly depend on different speakers or different syllables.

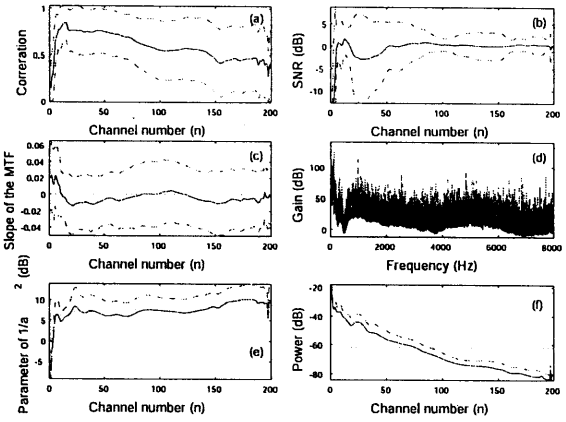

Figure 3: Results of analysis for dataset at point 2 in the same format as Fig. 2

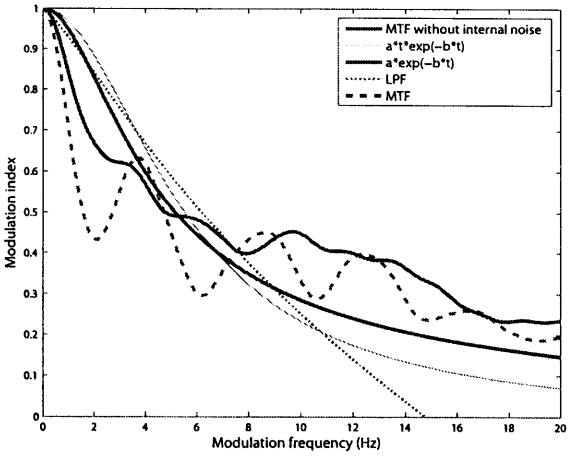

Figure 4: Comparison between MTF derived from AC/BC speech database and three fitted models

3.4 Functional modeling of MTF

In our previous method [4], MTF was represented as an exponential model. However, it was not evident whether MTF could be represented with an exponential curve. We confirmed that MTF has low-pass characteristics from the results of analysis. We modeled MTF by fitting three low-pass functions, i.e., exponential curve $e_h(t) = a_n t \exp(-b_n t)$, the model of MTF used in the previous method $e_h(t) = a_n \exp(-b_n t)$, and the $e_h(t) =$ IIR low-pass filter to the MTF derived from the database.

Figure 4 plots the results of comparing the MTF derived from the database and the three fitted functions. The shape of the MTF seems to be rippled. Because MTF without internal noise did not fluctuate, the influence of internal noise might ripple the shape of MTF. The model, $a_n \exp(-b_n t)$, was especially suitable in this comparison.

Figure 5 shows the mean and the standard deviation for the

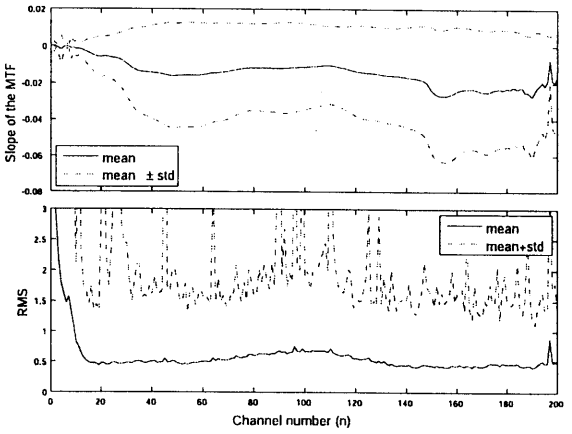

Figure 5: Results of analysis for MTF obtained by fitting model: Slope of regression line of fitted model (top) and RMS difference between model and MTF (bottom)

results of analysis using $a_n \exp(-b_n t)$. The top panel indicates the slope of the regression line of the model that was fitted by MTF derived from the database. If the value of the slope is zero, MTF does not influence BC speech. The bottom panel indicates the root mean squared (RMS) differences between the model and the MTF. As the RMS difference between MTF and $a_n \exp(-b_n t)$ was the smallest for all models, we assumed that the MTF relation, $e_h(t)$, could be represented by the model as $e_h(t) = a_n \exp(-b_n t)$. Then, we can determine inverse filtering as

$$E_h^{-1}(z) = \frac{1}{a_n^2}\left\{1 - \exp\left(-\frac{2b_n}{f_s}\right) z^{-1}\right\} \qquad (6)$$

where f_s is the sampling frequency of 16 kHz.

4. Method of MTF-Based Blind Restoration

Our previous method [4] needed information on AC speech to determine the parameters of the MTF model and to set the conditions for restoration. We improved our previous method in two respects using the results from analysis.

The first improvement was how we determined the two parameters of the model, a_n and b_n, to restore BC speech. From the results of analysis, parameter a_n can be approximated by a regression curve and this curve only depends on the measurement point. We can determine parameter a_n without AC speech to study each measurement of the regression curve.

To estimate parameter b_n, we used the method proposed by Hiramatsu and Unoki [9], which was originally introduced to estimate the reverberation time based on the MTF concept. We utilized this method because our research and theirs were based on exactly the same model. b_n can be estimated as

$$\hat{b}_n = \arg\min_{b_n} \left(\left| \hat{E}_y(0) \cdot |M(f_{dm}, b_n)| / \hat{E}_{y_n}(f_{dm}) \right| \right) \qquad (7)$$

where f_{dm} is the dominant frequency of the BC power envelope and $\hat{E}_y(\cdot)$ is the power envelope of restored speech.

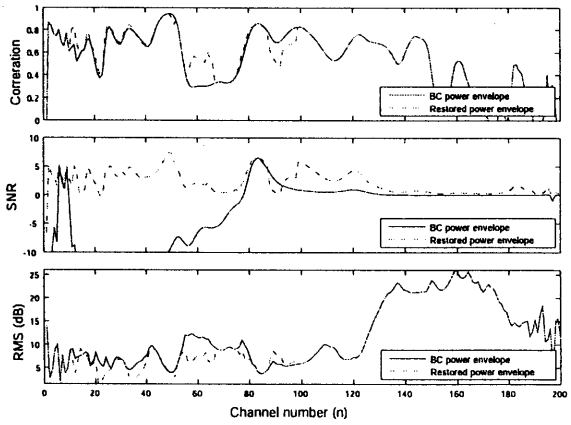

Figure 6: Improved Correlation, SNR and RMS of MTF

Our previous method restored BC speech when the correlation between AC and BC power envelopes was not over 0.8 and the relative power of AC power envelope was over -20 dB. The second improvement was that we changed these conditions so that the method restored BC speech when the relative power of the BC power envelope was over -40 dB to restore BC speech without using AC speech. The reason we set such conditions is that internal noise in BC speech appears when the relative power of the BC power envelope is not over -40 dB. Because internal noise increased the power of the DC of MTF, the method could not estimate the true value of parameter b_n. We therefore propose an MTF-based method for blindly restoring BC speech based on these results.

5. Evaluations

We carried out simulations to evaluate the new method using a subset of the AC/BC speech database. The correlation and SNR of the power envelopes for AC and the restored speech signals or the power envelopes for AC and BC speech signals were used to evaluate the improved restoration in the power envelopes. The perceptual evaluation of sound quality (PESQ) [10] that was recommended by ITU-T P. 862, and log spectral distortion (LSD) were used to evaluate the improved speech quality. The RMS of MTF was used to evaluate the improved intelligibility of restored speech. The RMS of MTF means the RMS difference between MTF that includes attenuation and MTF that does not include attenuation. If all modulation frequencies of MTF are 0 dB, MTF does not influence BC speech. Therefore, the closer the value of the RMS of MTF is to 0, the better the improvement in MTF.

Figure 6 shows an example of the SNR, correlation, and RMS of MTF. The solid lines indicate the results for BC speech and the dashed lines indicate the results for speech restored with the proposed method using parameters derived from the regression curve. These results demonstrate that the power envelopes can be adequately restored by the proposed approach. The LSD value decreased by 1.5 dB and the PESQ value increased by 0.7 points. These results revealed that sound quality can be improved with the new approach. The results of evaluations demonstrated that the new method could effectively and blindly restore BC speech.

6. Conclusion

We analyzed the characteristics of AC and BC power envelopes with an AC/BC speech database. As a result, we found that a characteristic of MTF was the low-pass filtering and each measurement point of power ratio could be approximated by a regression curve. We then modeled the MTF as $a_n \exp(-b_n t)$ and proposed methods of determining the parameters for the model without AC speech. We consequently proposed the MTF-based method of blind restoration. Finally, we evaluated the new approach and found that it could effectively and blindly restore BC speech.

We next intend to carry out comprehensive evaluations and assess how well the proposed method performs in these.

Acknowledgments

This work was supported by a Grant-in-Aid by the YAZAKI Memorial Foundation for Science and Technology It was also partially supported by the Strategic Information and COmmunications R&D Promotion ProgrammE (SCOPE) (071705001) of the Ministry of Internal Affairs and Communications (MIC), Japan.

References

[1] S. Kitamori and M. Takizawa: An analysis of bone conducted speech signal by articulation tests, IEICE Trans., Vol. J72-A, No. 11, pp. 1764–1771, 1989.

[2] S. Ishimitsu, H. Kitakaze, Y. Tsuchibushi H. Yanagawa and M. Fukushima,: A noise-robust speech recognition system making use of body-conducted signals, Acoust. Sci. &,Tech., Vol. 25, No. 2, pp. 166–169, 2004.

[3] T. Tamiya and T. Shimamura: Reconstruction filter design for bone-conducted speech, Proc. ICSLP2004, pp. 1085–1088, 2004.

[4] T. T. Vu, K. Kimura, M. Unoki and M. Akagi: A study on restoration of bone-conducted speech with MTF-based and LP-based models, J. Signal Processing, Vol. 10, No. 6, pp. 407–417, 2006.

[5] T. T. Vu, G. Seide, M. Unoki and M. Akagi: Method of LP-based blind restoration for improving intelligibility of bone-conducted speech, Proc. Interspeech2007, pp. 966–969, 2007.

[6] T. Houtgast and H. J. M. Steeneken: The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility, Acustica, Vol. 28, pp. 66–73, 1973.

[7] M. Drullman: Temporal envelope and fine structure cues for speech intelligibility, J. Acoust. Soc. Am., Vol. 97, pp. 585–592, 1995.

[8] Database for speech intelligibility testing using Japanese word lists. NTT–AT, 2003.

[9] S. Hiramatsu and M. Unoki: A Study on the Blind Estimation of Reverberation Time in Room Acoustics, J. Signal Processing, Vol. 12, No. 4, pp. 323–326. 2008.

[10] H. Yi and C. L. Philipos, :Evaluation of objective measures for speech enhancement, Interspeech2006, pp. 1447–1450, 2006.